

Swiss Finance Institute
Research Paper Series N°08 – 48

Preferences for Truthfulness: Heterogeneity Among and Within Individuals

Rajna GIBSON
University of Geneva and Swiss Finance Institute

Carmen TANNER
University of Zurich

Alexander F. WAGNER
University of Zurich, Swiss Finance Institute, Harvard University, and CEPR

swiss:finance:institute

 FINRISK

National Centre of Competence in Research
Financial Valuation and Risk Management

Established at the initiative of the Swiss Bankers' Association, the Swiss Finance Institute is a private foundation funded by the Swiss banks and Swiss Stock Exchange. It merges 3 existing foundations: the International Center FAME, the Swiss Banking School and the Stiftung "Banking and Finance" in Zurich. With its university partners, the Swiss Finance Institute pursues the objective of forming a competence center in banking and finance commensurate to the importance of the Swiss financial center. It will be active in research, doctoral training and executive education while also proposing activities fostering interactions between academia and the industry. The Swiss Finance Institute supports and promotes promising research projects in selected subject areas. It develops its activity in complete symbiosis with the NCCR FinRisk.

The National Centre of Competence in Research "Financial Valuation and Risk Management" (FinRisk) was launched in 2001 by the Swiss National Science Foundation (SNSF). FinRisk constitutes an academic forum that fosters cutting-edge finance research, education of highly qualified finance specialists at the doctoral level and knowledge transfer between finance academics and practitioners. It is managed from the University of Zurich and includes various academic institutions from Geneva, Lausanne, Lugano, St.Gallen and Zurich. For more information see www.nccr-finrisk.ch.

This paper can be downloaded without charge from the Swiss Finance Institute Research Paper Series hosted on the Social Science Research Network electronic library at:

<http://ssrn.com/abstract=1323283>

Preferences for Truthfulness: Heterogeneity Among and Within Individuals [♦]

Rajna Gibson ^{*}, Carmen Tanner ^{**}, and Alexander F. Wagner ^{***}

April 10, 2012

Forthcoming, *American Economic Review*

Abstract

We conduct an experiment assessing the extent to which people trade off the economic costs of truthfulness against the intrinsic costs of lying. The results allow us to reject a type-based model. People's preferences for truthfulness do not identify them as only either "economic types" (who care only about consequences) or "ethical types" (who care only about process). Instead, we find that preferences for truthfulness are heterogeneous among individuals. Moreover, when examining possible sources of intrinsic costs of lying and their interplay with economic costs of truthfulness, we find that preferences for truthfulness are also heterogeneous within individuals.

Keywords: Truthfulness, heterogeneous preferences, type-based models, costs of lying, honesty, incentives, earnings management

JEL-Codes: A13, C91, G30, M14

[♦] We thank the Research Priority Program Finance and Financial Markets of the University of Zurich, the NCCR FINRISK, the Swiss Finance Institute, and the Swiss National Science Foundation (Grant #PP001-102845) for financial support. Nicolas Berkowitsch provided excellent research assistance. We thank Marcel Morf, who programmed the first version of the code for the experiments. We thank the Editor (Joel Sobel) and three anonymous referees for excellent remarks that have greatly improved the paper. Rainer Winkelmann, who went far beyond the call of duty in helping us clarify the empirical framework and its relation to our preferred preference specification, deserves our special thanks. We also thank Peter Bossaerts, Chiddi Chidambaram, Alain Cohn, François Degeorge, Ernst Fehr, Gerlinde Fellner, Laurent Frésard, Michelle Sovinsky, Susanne Neckermann, Pawel Polak, Bernard Raffournier, Jean-Charles Rochet, Meir Statman, and Mei Wang for fruitful discussions and comments. Useful comments were provided by seminar participants at the European Finance Association meetings, the International Conference on Psychology, the Conference of the Society of Judgment and Decision Making, the European Association of Experimental Social Psychology, la Universidad Carlos III, Fudan University, the University of Fribourg, Northwestern University, Rutgers University, the University of St. Gallen, and the University of Zurich. Part of this work was conducted while Gibson was visiting UCLA.

^{*} Swiss Finance Institute -- University of Geneva. Mail: University of Geneva, 40 Bd Pont d'Arve, CH-1211 Genève 4, Switzerland, +41 (022) 379 8983, Rajna.Gibson@unige.ch

^{**} University of Zurich. Mail: University of Zurich, Department of Banking and Finance, Plattenstrasse 32, CH-8032 Zurich, Switzerland, +41 (044) 634 4016, carmen.tanner@bf.uzh.ch

^{***} Swiss Finance Institute - University of Zurich, Harvard University, and CEPR. Mail: University of Zurich, Department of Banking and Finance, Plattenstrasse 14, CH-8032, Zurich, Switzerland, +41 (044) 634 3963, alexander.wagner@bf.uzh.ch

Standard economic models of self-interested utility maximization, which emphasize the role of consequences in determining agents' actions, predict a grim inexorability to all economic systems. These models are based on an assessment of humans as self-interested agents who behave dishonestly for cogent reasons. These hypothetical persons prioritize the outcomes of their actions and only forgo materially beneficial lying if strategic or reputational considerations arise. Some researchers, such as Amar Bhidé and Howard H. Stevenson (1990), assert that these reputational forces are often weak, implying that honesty simply does not seem to pay.

Examples of disastrous dishonesty based on such self-interest abound in the corporate world. Deliberate deception has augmented the economic effects of regulatory failure, of a deteriorating macro-economy, and of inadequate models, in, for example, the subprime crisis.

Yet, truthfulness also appears to prosper in society. Whistleblowers often jeopardize their careers and friendships when they truthfully reveal the wrongdoing of their companies. Some CEOs are regarded as particularly virtuous (Linda K. Treviño and Michael E. Brown 2004). Numerous journalists risk their lives to report the truth about political repression, economic crimes, and human rights violations.

To explain otherwise puzzling behavior both in the field and in experiments, several authors have proposed the idea that some people experience intrinsic costs when they lie. For example, in a cheap-talk sender-receiver game, Uri Gneezy (2005) found that many subjects told the truth.¹ Of various possible explanations for this result, he inferred that the most plausible was that “people have non-consequential preferences in which they treat the same monetary outcome differently, depending on the process that leads up to it” (p. 392). Moreover, in Gneezy's interpretation, “different people weigh these preferences differently” (ibid.). That is, a model in which agents exhibit (continuously) *heterogeneous preferences for truthfulness* could explain his data.

Because Gneezy's experiment was set in a strategic context, social preferences may also have been active. Therefore, Gneezy also emphasized the joint relevance of process-

¹ Similar results on truth-telling have been obtained in other studies (John H. Evans, R. Lynn Hannan, Ranjani Krishnan, and Donald V. Moser 2001; Santiago Sánchez-Pagés and Marc Vorsatz 2007). Only a few researchers, such as Stanley Baiman and Barry L. Lewis (1989), have found that people will lie even for just a tiny monetary payoff. See the edited volume by Paul J. Zak (2008) for numerous additional examples.

dependent preferences and of consequences to oneself and to others.² But it is precisely because both lying aversion and social preferences operated in his experiment that the two channels were difficult to isolate. Indeed, Sjaak Hurkens and Navin Kartik (2009, p. 180) showed that Gneezy's (2005) empirical observations were consistent with the "hypothesis that people are one of two kinds: either a person will never lie, or a person will lie whenever she prefers the outcome obtained by lying over the outcome obtained by telling the truth." Based on this existing evidence, it is, therefore, possible that the world is populated, in the spirit of *type-based models* such as that of Kenneth Koford and Mark Penno (1992), by exactly two fixed types: "economic types" and "ethical types" (in Gneezy's terminology). Alternatively, these two types can, respectively, be characterized as consequentialists (who care about consequences to themselves and to others, but not about the process by which these consequences are achieved) and as non-consequentialists (who care only about the process, but not about consequences).

The two-type-based model and the model with heterogeneous preferences for truthfulness lead to very different implications, particularly for agent selection and incentive design. Therefore, it is important to determine which of these two models offers a more accurate description of reality.

To address this question, we conducted a decision-theoretic laboratory experiment in which each participant was placed in the situation of a CEO who had to announce his/her firm's earnings to a passive market. The participants were informed of the true level of earnings. They were also told that falsely reporting higher earnings was legal and would lead to higher actual payoffs than honestly announcing the lower earnings. We considered that economic types would always lie in our experiment because truthfulness was designed to be economically costly; we considered that ethical types would always tell the truth. If, by contrast, individuals varied continuously in the extent to which they were driven by preferences for truthfulness, they would trade off the economic costs of truthfulness with the costs of lying; those with intermediately strong preferences for truthfulness would exhibit the most changes in behavior as economic costs changed.

² A long-standing literature considers the role of preferences that depend on process and/or on consequences for others; it recognizes that people do not necessarily maximize utility according to the material consequences of their actions. For example, Matthew Rabin (1995) demonstrated how fairness considerations can explain why people are willing to reward or punish others even when this requires a sacrifice of their own well-being.

The simplicity of our experimental setup--involving a decision-making situation with no counterparties--allowed us to isolate motivations for truthtelling that are non-strategic and not driven by social preferences; it permitted us to sidestep issues that occur in strategic contexts. Moreover, in our experiment, we observed individuals' behaviors. This setup enabled us to provide evidence, stronger than that developed in existing works, regarding heterogeneity in preferences for truthfulness.

We observed that, in a situation where the standard economic model predicts that everybody will lie, 32% of the participants chose not to do so, thus forgoing a larger variable compensation. Importantly, the aggregate percentage of truthtellers decreased as the costs of truthfulness increased. Our individual-level regressions imply that the marginal effect of a cost increase on the probability of an individual's telling the truth is significantly negative, even after controlling for various demographic and psychological factors. These results are at odds with the type-based model but are consistent with a model that posits heterogeneous preferences for truthfulness.

Our primary contribution, therefore, is to provide evidence for the notion that people occupy a spectrum of preferences for truthfulness rather than only two opposite positions. As a secondary contribution, we examine potential sources of the heterogeneity in preferences for truthfulness. Tendencies towards impression management and self-deception offer no explanatory power; however, one measure of one source of intrinsic costs of lying, an index of "protected values of truthfulness," seems to organize the data well. We also find substantial evidence of non-separability between this measure of intrinsic costs of lying and economic costs of truthfulness in the utility function. In other words, total preferences for truthfulness not only display heterogeneity among, but also within, individuals. We do not have adequate measures of other possible sources of intrinsic costs of lying, including, in particular, measures of expressive preferences. Therefore, we acknowledge that other preference formulations could potentially explain our empirical evidence.

Section I presents the basic tradeoff and the hypotheses. Section II describes the experiment. Section III discusses the main results. Section IV explores possible sources of intrinsic costs of lying and their interaction with economic costs of truthfulness. Section V concludes.

I. The tradeoff

Consider an agent who decides whether to tell the truth, $T=1$, or to lie, $T=0$. Lying, the agent receives a certain income m . There are *Economic Costs Of Stating the Truth*, for which we use the term *ECOST*. The agent receives funds $m-ECOST$ when he tells the truth. We model *preferences for truthfulness* by positing that the agent also experiences *total costs of lying*, C_i . (For the moment, C_i is given. We discuss in Section IV how these total costs of lying may arise from the interplay between the intrinsic costs of lying and the extrinsic economic costs of truthfulness.) If types are continuous, C_i can take on any value, positive or negative. By contrast, in the two-type model, there are only “ethical types” who have $C_i = \infty$ and “economic types” who have $C_i = 0$. Let the global utility function be defined as

$$V_i(T) = \begin{cases} b(m - ECOST) & \text{if } T = 1 \\ bm - C_i & \text{if } T = 0 \end{cases} \quad (1)$$

where T is the choice variable.³ For simplicity, and because wealth effects are unlikely in our experiment, we assume the agent has a constant marginal utility of money $b > 0$. We also assume that all participants have the same b .⁴

The difference between the utilities of truthtelling and of lying is given by

$$Y_i^* = C_i - bECOST. \quad (2)$$

An individual exhibits truthfulness when $Y_i^* > 0$. This implies that truthfulness can, in this framework, only arise as optimal behavior if there is a positive total cost of lying. While social preferences are known to contribute to behavior (e.g., Ernst Fehr and Urs Fischbacher 2002, 2003), our experiment is designed to eliminate any role for altruism, reciprocity, guilt aversion (Gary Charness and Martin Dufwenberg 2006), and related factors, as well as any role for strategic concerns that might arise with repeated interaction.

³ Truthfulness here is a matter of preference. Alternatively, we could posit a constraint involving a need to maintain a minimum level of truth-telling. Within this simple context, the two formulations are identical. Rabin (1995) showed that moral preferences and moral constraints can result in different incentives for information collection.

⁴ It is standard to assume that, abstracting from the preference feature of interest (for example, inequality aversion), all participants have equal marginal utility of money. See, for example, Ernst Fehr and Klaus Schmidt (1999).

Consider now a population of individual decision-makers (whose distribution of C_i is not known), each of them facing various economic costs of truthfulness. A type-based model, such as that of Koford and Penno (1992), implies that ethical types, with their overwhelming preferences for truthfulness, would always choose $T=1$, and this choice would be invariant to $ECOST$. Conversely, economic types would always lie when profitable. (At $ECOST = 0$, they would perceive no advantage or disadvantage to either telling the truth or lying; but at all other levels of economic costs of truthfulness, the utility difference Y_i^* would be negative.) Aggregating across the population of individuals, this implies the following hypothesis:

Hypothesis TYP (Type-based model): *The fraction of the population telling the truth remains constant across varying economic costs of truthfulness.*

By contrast, in the model based on heterogeneous preferences for truthfulness, where C_i varies continuously throughout the population, varying economic costs would lead some individuals with intermediate total lying costs to change their behavior. Higher economic costs of truthfulness would then make it less likely that an individual would tell the truth. Thus, we have the alternative hypothesis reflecting Gneezy's conjecture:

Hypothesis HET (Model based on heterogeneous preferences for truthfulness): *The fraction of the population telling the truth varies with economic costs of truthfulness.*

In Section III, we test these two hypotheses using aggregate behavioral data. We also specify an empirical model for individual choice to test the corresponding underlying predictions regarding the marginal effect of economic costs of truthfulness on individual choice.

II. Experimental method⁵

We are interested in situations requiring a choice between telling the truth and telling a lie, in which the former decision involves an economic sacrifice. As our context, we chose accounting earnings management (henceforth called “earnings management”).⁶ This situation illuminates a real-life conflict: management’s variable compensation is frequently tied to stock price performance, which in turn often hinges on earnings announcements. We envisioned a framework in which earnings management would be understood to be legal (for example, within GAAP rules) although explicitly self-interested and dishonest—a decision-making problem focused exclusively on the managerial choice. We required the recipient (the market, played by the computer) to accept passively all financial statements. The advantage of this approach is that, due to the absence of strategic interactions, we have been able to isolate, at least better than in the real world, factors influencing individuals’ choices, without monitoring the participants’ thoughts regarding the behavior of other players.⁷

A. *Participants and procedure*

A total of 261 participants (median age: 23 years) took part in this online experiment. We recruited participants from undergraduate classes at the University of Zurich (Switzerland). 50% of the participants were economics and finance students, 40% psychology students, and 10% students from other fields. 42% were women and 58% were men (distributed

⁵ The full set of instructions is available in a Supplementary Appendix. The experiment included tasks whose content and results are not described here for space reasons.

⁶ Accounting earnings management occurs “when managers use judgment in financial reporting and in structuring transactions to alter financial reports to either mislead some stakeholders about the underlying economic performance of the company or to influence contractual outcomes that depend on reported accounting numbers” (Paul M. Healy and James M. Wahlen 1999). Accounting earnings management can be viewed as a form of lying, which is defined as “a statement that one knows to be false” (Steven L. Grover 2005). Mark W. Nelson, John A. Elliott, and Robin L. Tarpley (2003) provide many examples collected from auditors.

⁷ In particular, despite the intuitive appeal and real-world relevance of the strategic games employed in Gneezy (2005) and in Hurkens and Kartik (2009), these games come with some interpretational challenges. For instance, in sender-receiver games, even telling the truth can be deceptive, because the sender may hope that the receiver will not believe the true message that is sent (Matthias Sutter 2009). Additionally, Julian Rode (2010) found that decision makers were significantly less trusting in a competitive context than in a cooperative context.

across the fields). All participants were told at the outset that anonymity was ensured.⁸ They were first asked to respond to a few demographic questions and to read some basic instructions. They were informed that they would individually receive a payment, CHF 8, for their completed participation in the study, and an additional payment that depended on their decisions. After having demonstrated their understanding of the (unlabelled) tasks and of the rules of the experiment, the participants completed, in randomized orders, the three main parts of the experiment: 1) the truthtelling task, 2) the effort task, and 3) the measurement of various controls and potential proxies for intrinsic costs of lying. Finally, all the participants were paid. For simplicity, we describe the procedure for one of the randomized orders of tasks.

1) In the truthtelling task, each participant was placed in the situation of a CEO who had to announce earnings per share for the previous quarter. The participants were told that the variable component of their salaries would depend on the earnings they announced. They were also told that the market currently anticipated the announcement of 35 cents per share as earnings, but that the true earnings were 31 cents per share. The participants were told that they could announce earnings of 35 cents per share while remaining within legal accounting limits,⁹ and that the decision would be solely theirs. They were also informed that they would be paid an amount based on the CEO compensation (according to their decisions). This additional experimental payoff would be converted into real money at the rate of CHF 100,000 = CHF 0.5. Importantly, participants earned less when choosing to tell the truth.

The participants were then told they would have to announce their financial statements that day. The truthtelling task questionnaire follows:

⁸ At the time of the experiment, the exchange rate was about US \$1 = CHF 1.15. Most participants received payment one week after the experiment. For this purpose, each participant received, before the experiment, a code, based on which the experimenter prepared an envelope containing the earnings. Participants received the sealed envelopes by indicating their personal codes. It is, therefore, unlikely that a desire to appear honest affected the participants' behavior systematically. Dan Ariely, Anat Bracha, and Stephan Meier (2009) document how publicly displayed monetary incentives can be less effective in promoting pro-social behavior than privately displayed incentives.

⁹ Therefore, risk preferences of individuals did not matter, as their choices were not based on the trade-off between the expected benefits and costs of committing a crime.

Which earnings will you announce?

- ☐ 31 cents per share -- In this case, your compensation will be CHF 60,000 (CHF 0.30).
- ☐ 35 cents per share -- In this case, your compensation will be CHF 300,000 (CHF 1.50).

- ☐ 31 cents per share -- In this case, your compensation will be CHF 120,000 (CHF 0.60).
- ☐ 35 cents per share -- In this case, your compensation will be CHF 300,000 (CHF 1.50).

- ☐ 31 cents per share -- In this case, your compensation will be CHF 180,000 (CHF 0.90).
- ☐ 35 cents per share -- In this case, your compensation will be CHF 300,000 (CHF 1.50).

- ☐ 31 cents per share -- In this case, your compensation will be CHF 240,000 (CHF 1.20).
- ☐ 35 cents per share -- In this case, your compensation will be CHF 300,000 (CHF 1.50).

- ☐ 31 cents per share -- In this case, your compensation will be CHF 300,000 (CHF 1.50).
- ☐ 35 cents per share -- In this case, your compensation will be CHF 300,000 (CHF 1.50).

A few questions served as a manipulation check to verify that participants distinguished between the 31 and 35 cent options. The participants were asked, using a 5-point scale ranging from -2 to +2, the extent to which they judged announcing 31 cents as *dishonest* vs. *honest*, *manipulative* vs. *not manipulative*, *short-term-oriented* vs. *long-term-oriented*, and *associated with personal benefits* vs. *associated with personal costs*. The same was also done for the 35-cent announcement option.

2) Participants engaged in a simple calculation (effort) task.

3) We then measured, as potential sources of intrinsic costs of lying (a term we introduce formally in Section IV), their tendencies towards impression management and self-deception, and their levels of protected values. Moreover, we also measured their altruistic concerns.

After the experiment, the participants anonymously received their payments of CHF 8 plus their earnings. The average total payment was slightly less than CHF 30.5.¹⁰

¹⁰ This amount includes payment for other tasks in the full experiment, as described in the Supplementary Appendix.

B. Variables of interest

TRUTHFUL CHOICE. This represented the dependent variable in the truthtelling task, coded as a binary variable that took on the value of 1 if a participant chose to announce earnings of 31 cents (the honest option), and the value of 0 if a participant announced 35 cents (the dishonest option).

ECOST. This was a within-participants variation. Economic costs of truthfulness derived from the amount of money a participant forfeited by announcing 31 cents. The *ECOST* variable took on values from CHF 0 to CHF 1.20 (= 1.50 – 0.30), in increments of 30 cents.

IMPRESSION MANAGEMENT and SELF DECEPTION. Using the standard Deception Scales (PDS) of Delroy Paulhus (1984) in the German version of Jochen Musch, Robbi Brockhaus, and Arndt Bröder (2002), we measured individuals' tendencies to give socially desirable responses. These tendencies come in two distinct forms: a tendency to deceive others (impression management) and a tendency to deceive oneself (self-deception). Both are expected to be positively related to intrinsic costs of lying. Accordingly, we coded two variables, *EXTDECEIT* and *SELFDECEIT*. We scaled the measures to be between 0 and 1. Participants who exhibited more socially acceptable responses scored higher on both scales.

PROTECTED VALUES (PV). The extent to which participants held truthfulness as a protected value and, therefore, felt committed to truthtelling was another source of intrinsic costs of lying. To measure this source, we used an index developed by Carmen Tanner, Bettina Ryf, and Martin Hanselmann (2009), the details of which are available in the Supplementary Appendix. This index took on a value between 0 (for an individual with no protected values) and 6 (for an individual with maximum protected values). The internal consistency of this scale, as assessed by Cronbach's α , was very satisfactory ($\alpha = 0.86$).¹¹

ALTRUISTIC CONCERNS. We asked participants the extent to which they believed that announcing 31 cents (or 35 cents) had consequences for other stakeholders (-2 = *hurting other stakeholders* to +2 = *not hurting other stakeholders*). Of course, within the strict confines of the experiment, there were no such consequences. Nonetheless, this

¹¹ Cronbach's Alpha is a measure of the reliability and the internal consistency of an instrument. The measure ranges from 0 to 1 and will generally increase when the correlations between the items increase.

variable was a relevant control for any altruistic preferences or fairness concerns of the participants which might confound our inferences. Answers to this question were coded as the variable *35HURTS*.

DEMOGRAPHIC CONTROL VARIABLES. *SEX* was equal to 1 for female participants and to 0 for male participants. *AGE* was equal to each participant's age in completed years. *PSYCHOLOGY* was equal to 1 for psychology students ("psychologists") and to 0 otherwise. *OTHER* was equal to 1 for participants from fields other than psychology and economics and to 0 otherwise. *ECONOMICS* was the omitted category.

III. Main results

A. Descriptive evidence

We first confirm, through a manipulation check, that the participants generally understood the announcement of 31 cents to be the honest, non-manipulative action that would lead to a personal loss, while the opposite was true of the announcement of 35 cents. (See Table I.)

TABLE I ABOUT HERE

Table II allows a first look at the choices the participants made in the experiment. In approximately 42% of cases (32% when omitting the free-truth situation), participants chose to announce low earnings, that is, chose not to engage in earnings management. By telling the truth, those participants opted to suffer, on average, effective monetary losses of 11% of the maximum total amount they could have earned in the truth-telling task or 27% of the variable amount they could have earned above the guaranteed payout.

TABLE II ABOUT HERE

The fact that a large proportion of the participants reported the truth, even when the conditions opposed it, is consistent with the notion that many individuals have positive total costs of lying. By contrast, this finding is inconsistent with the standard economic model.

B. Reactions to economic costs of truthfulness

In this section, our primary goal is to test the implications of the type-based model, which posits that the (participant) population consisted only of “economic types” and “ethical types,” against the implications of the model based on heterogeneous preferences for truthfulness.

Table II, showing aggregate data, reveals substantial variation in the participants’ responses as the economic costs of truthfulness changed: with higher economic costs, the percentage of participants telling the truth was lower. This is inconsistent with the type-based model’s prediction in Hypothesis TYP, according to which the fraction of participants who told the truth would have remained constant. Formally, a χ^2 -test strongly rejects the hypothesis that there is a fixed fraction of “ethical” types who always tell the truth and a fixed fraction of “economic types” who always lie, with nobody differing from these two types. (This is true not only of the specific version postulated by Hurkens and Kartik (2009), where exactly half of the population always tell the truth and the other half always lie, but also for any other fraction between 0 and 1.) When there was no economic cost of truthfulness, 18% of the participants still chose the earnings-management solution. This can be explained by recognizing that the model based on heterogeneous preferences for truthfulness allows agents to have a negative total cost of lying. (The manipulation check confirms that this group of people perceived 35 cents as the less honest option.)

To investigate statistically the influence of the economic costs of truthfulness on individual behavior, we estimate a discrete choice / random utility model (e.g., Gary King 1998; Jeffrey M. Wooldridge 2006). From Equation (2), adding a stochastic error and rearranging terms, each participant i ’s ($i=1, \dots, 261$) latent utility difference between truth-telling and lying at direct economic $ECOST_j$ is given by

$$Y_{ij}^* = C_i - bECOST_j + \varepsilon_{ij}. \quad (3)$$

Under utility maximization, an observed realization of *TRUTHFUL CHOICE*, T_{ij} , is related to Y_{ij}^* by the following mechanism:

$$T_{ij} = \begin{cases} 1 & \text{if } Y_{ij}^* \geq 0 \\ 0 & \text{if } Y_{ij}^* < 0. \end{cases} \quad (4)$$

In line with standard practice, we assume that ε is independent of the explanatory variables \mathbf{X} . By assuming that ε has the logistic distribution, one obtains the logit model, which is the main specification on which we focus. After relabeling and combining coefficients,

$$\Pr(T_{ij} = 1|\mathbf{X}) = \Lambda[\beta_0 + \beta_E ECOST_j], \quad (5)$$

where $\Lambda(\bullet)$ is the logistic cumulative distribution function.¹² The coefficient vector is estimated by maximum likelihood. If $\hat{\beta}_E < 0$, participants react negatively overall to economic costs. The coefficient $\hat{\beta}_0$ here gives the average C_i in the data for zero economic costs of truthfulness. (In Section IV, we explore sources of variation in C_i and we discuss that, besides the direct effect of reducing the attractiveness of truthfulness, *ECOST* may have an indirect effect through the total costs of lying.) The standard errors correct for possible serial correlation and heteroskedasticity by clustering at the individual level. (Recall that participants went through all five economic cost situations.)

Column (1) of Table III shows the results of this analysis. Consistent with the fact that many individuals did, in fact, tell the truth, the constant term is positive. Importantly, *ECOST* is a highly significant determinant of the relative attractiveness of truthfulness and lying for an individual. Indeed, the implied marginal effect of *ECOST* is powerful: A 30-percent increase in *ECOST* was associated with a 16.9% decrease in truth-telling.¹³ Together with the observations made in Table II, this finding supports Hypothesis HET.

TABLE III ABOUT HERE

Column 2 of Table III adds individual-level controls. The main result for *ECOST* remains unchanged. We observe some interesting additional findings. First, women appeared to be more likely to tell the truth, as did students in fields outside of psychology and economics. Second, given the decision-theoretic nature of the experiment, altruistic and distributional concerns, as well as attempts to live up to others' expectations so as to avoid guilt, should

¹² If ε is normally distributed, one obtains the probit model. As is typical in econometric applications, the two models yield virtually identical inferences.

¹³ In the real world, managers are indeed faced with substantial cross-sectional and time-series variation in the economic cost of truth-telling. Our results are consistent with findings by Daniel Bergstresser and Thomas Philippon (2006), who showed that the use of discretionary accruals to manipulate reported earnings was more pronounced at firms where CEO compensation depended more on the stock price.

not have affected behavior in this experiment. Yet, we note that *35HURTS* does enter significantly in this baseline regression. This suggests one of two possibilities. Either participants' altruistic concerns drove behavior, or *35HURTS* was correlated with some general differences in preferences that were in turn correlated with intrinsic costs of lying and thus reflected variation in a variable omitted in Column 2. As Section IV shows, the latter explanation is corroborated by the data.

The conclusions we draw from these main results are simple but important. Hurkens and Kartik (2009) demonstrated that Gneezy's (2005) data would be consistent with a population of pure opportunists, who always lie, and pure ethical types, who always tell the truth (as in Koford and Penno 1992). However, the present evidence of changeability in truth-telling behavior and of significant sensitivity to economic costs associated with truthfulness rejects Hypothesis TYP and is in direct contrast to the implications of a type-based model.

IV. Sources of heterogeneity in total costs of lying

We have established that the participants in our experiment showed more variation in their total costs of lying than if they had belonged to one of just two fixed extreme types. In this section, we expand on these findings by considering various potential sources of the variation among individuals in total costs of lying. Moreover, this extension allows us to provide further evidence refuting the type-based model; in particular, we document that there is also heterogeneity *within* individuals (across situations) in total costs of lying.

A. Enhanced model

We consider two constituent sources of heterogeneity in total costs of lying. First, these costs are driven by individuals' *Intrinsic Costs Of Lying*, for which we use the term $ICOL_i$. Second, we also allow the situation, that is, the economic costs of truthfulness or the extrinsic incentives for lying, $ECOST_j$, to influence the total costs of lying. Adjusting notation, we posit that total costs of lying may vary both among and within individuals, and we now write $C_{ij} = C_{ij}(ICOL_i, ECOST_j)$. Since all participants encountered the same

ECOST situations, the evidence on heterogeneous total costs of lying provided in Section III necessarily implies that there is heterogeneity in $ICOL_i$. Next, the intrinsic costs of lying and the economic incentives for lying (economic costs of truthfulness) may enter C_{ij} separably or non-separably. Indeed, whether intrinsic preferences and extrinsic incentives interact in determining total preferences for a certain action has implications that extend beyond the scope of the current study; see, for example, Samuel Bowles and Sandra Polanía Reyes (2011) for a discussion of how incentives to contribute to public goods may affect social preferences. In order to capture both possibilities, we consider, for parsimony, a simple parametric specification

$$C_{ij}(ICOL_i, ECOST_j) = \gamma_0 + \gamma_1 ICOL_i + \gamma_2 ECOST_j + \gamma_3 ICOL_i ECOST_j, \quad (6)$$

implying

$$\begin{aligned} Y_{ij}^* &= C_i - b ECOST_j + \varepsilon_{ij} = \\ &= \gamma_0 + \gamma_1 ICOL_i + (\gamma_2 - b) ECOST_j + \gamma_3 ICOL_i ECOST_j + \varepsilon_{ij} \end{aligned} \quad (7)$$

and, with the maintained distributional assumptions,

$$\Pr(T_{ij} = 1 | \mathbf{X}) = \Lambda[\beta_0 + \beta_I ICOL_i + \beta_E ECOST_j + \beta_{IE} ICOL_i ECOST_j], \quad (8)$$

where $\hat{\beta}_0, \hat{\beta}_I, \hat{\beta}_E$, and $\hat{\beta}_{IE}$ are the implied estimates for the model parameters $\gamma_0, \gamma_1, (\gamma_2 - b)$, and γ_3 , respectively.

B. Interpreting the model coefficient estimates

Naturally, C_{ij} is expected to be increasing in $ICOL_i$, so that we predict $\hat{\beta}_I > 0$. This coefficient thus allows us to test whether a candidate measure of $ICOL$ helps explain heterogeneity *among* individuals in total costs of lying. Moreover, specification (6) allows for two channels through which heterogeneity in total preferences for truthfulness *within* individuals enters. First, it seems reasonable to postulate that C_{ij} is increasing in $ECOST$ so that individuals associate higher monetary stakes with a stronger preference to tell the truth. Given that the parameters b and γ_2 are not identified by the present approach, $\hat{\beta}_E < 0$ only tells us that C_{ij} is not increasing in $ECOST$ at a rate greater than marginal utility b .

Second, heterogeneity of preferences within individuals for truthfulness can unambiguously be detected by considering the interaction term β_{IE} . Under the discrete-choice model's assumptions, a significant interaction term provides further evidence against the notion, posited by the type-based model, that there are two *fixed* types.¹⁴ Specifically, a positive interaction term arises if the intrinsic costs of lying are more important in determining total preferences for truthfulness when the stakes (*ECOST*) are higher. An equivalent interpretation (useful in settings where an agent can choose the size of the lie) is that individuals with stronger intrinsic costs of lying perceive “larger” lies, which yield larger economic benefits, as less attractive and will, thus, tell “smaller” lies. A negative interaction term instead arises if the source of the intrinsic costs of lying is relatively less influential at higher stakes. If the coefficient on the interaction is zero, the economic costs of truthtelling are perceived identically by all agents, regardless of the strength of their intrinsic costs of lying, $ICOL_i$. In that case, all agents' utilities would react identically to changes in the economic costs, even though heterogeneous $ICOL_i$ would imply that some would report the truth while others would lie at a given *ECOST*.¹⁵

¹⁴ Alternatively, if the true utility function has a separable form but the assumption of weak exogeneity of the error term ε_{ij} does not hold, then the interaction term in the model may serve as an instrument to correct for correlation between the explanatory variables ($ICOL_i$ and $ECOST_j$) and the error term ε_{ij} . The interaction term can serve as an instrument because it arises as one of the terms in the second-order Taylor-series expansion of the random utility function (with violated weak exogeneity). It captures jointly the effects of both explanatory variables. One anonymous Referee provided an example in which 90% of individuals make deterministic decisions based on an additively separable utility function, i.e., they tell the truth when $C_i = ICOL_i$ is greater than *ECOST*, and they lie when $ICOL_i$ is smaller than *ECOST*. 10% of individuals make decision errors; that is, they tell the truth although *ECOST* is greater than $ICOL_i$, and they lie even when $ICOL_i$ is greater than *ECOST*. In simulated data, the Referee showed that a logit regression (which is a misspecified model under the assumptions made) may yield a significant interaction term on *ECOST* and intrinsic costs of lying, even in this setting. In the Referee's example, the error term is not independent of the observable variables. Thus, the interaction term becomes significant in this setting because it is an instrumental variable for an omitted variable. We conclude that, even if the true utility function is separable, researchers may well want to use a non-separable reduced form because this specification is robust to the violation of the assumption of weak exogeneity of the error term, such as occurs in decision errors of the form proposed by the Referee.

¹⁵ When testing for non-separability (that is, for the significance of the interaction term), we consider coefficients, rather than marginal effects, from the logit regressions. Recall that marginal effects in a logit regression are given by $\Lambda(\beta'X)(1 - \Lambda(\beta'X))\beta$, where $\Lambda(\bullet)$ is the logistic cumulative distribution function giving the initial probability of truthfulness. Those with high (low) *ICOL* have high (low) initial probabilities of truthfulness. Thus, the highest marginal effects of *ECOST* on behavior are likely to be found in the middle range, and smaller marginal effects are likely to be found among those with high intrinsic costs. Analyzing coefficients instead allows us to consider the hypothetical case of participants who would display identical initial probabilities of reporting the truth.

C. Results

We first consider three possible sources of intrinsic costs of lying, $ICOL_i$, for which survey measures are available to us, and we then discuss other possible drivers of behavior. The descriptive statistics for *EXTDECEIT*, *SELFDECEIT*, and *PV* shown in Table IV indicate that there is wide variation in these three variables, suggesting that they could potentially explain the observed variation in truthtelling behavior. In Table III, we test whether this is the case. We allow each possible source to affect behavior both separably from economic costs and jointly by way of an interaction.

TABLE IV ABOUT HERE

First, participants may have developed an interest in impressing the experimenter by appearing honest and non-greedy (e.g. Urs Fischbacher and Franziska Heusi 2008); this would act like a preference for truthfulness. Given the design of the experiment, in which we took great care to make the responses anonymous, this is very unlikely to have occurred. Indeed, *EXTDECEIT* is not significant in any of the regressions, and neither is the interaction term with *ECOST*.

Second, it is possible that participants deceived or impressed themselves by making the “right” choices. However, *SELFDECEIT* is also not significant in any of the regressions, and neither is the interaction term with *ECOST*.

Third, we consider the possibility that moral values were a source of the intrinsic costs of lying. While many moral concepts are potentially relevant, we focus on protected values (*PV*). The literature that has developed the theory of these values emphasizes that protected values are non-consequentialist and induce a resistance to engaging in actions that would violate moral values, reducing the attractiveness of any financial gains obtained through such actions.¹⁶ That is, the economic costs of truthfulness matter less to those who hold stronger protected values of truthfulness; those people are trade-off resistant. This idea

¹⁶ See, for example, Jonathan Baron and Mark Spranca (1997); Philip E. Tetlock, Orie V. Kristel, S. Beth Elson, Melanie C. Green, and Jennifer S. Lerner (2000); and Carmen Tanner, Douglas L. Medin, and Rumen Iliev (2008). The source of protected values is modeled by Roland Bénabou and Jean Tirole (2011) as a need of agents to invest in their identity. For the strongest form of *PV*, “sacred” values and taboos, see in particular section V of their paper.

naturally translates into a functional form for C_i that is non-separable into intrinsic (moral) costs of lying and economic costs of truthfulness.

Column (3) of Table III shows that PV of truthfulness was a highly significant predictor of behavior in the experiment. A one-point increase in PV was associated with a 17.4% increase in the probability of truthtelling, holding the other variables at their means. In Column (4), we obtain a positive, significant coefficient on the interaction term between PV and $ECOST$. This is evidence that, conditional on the correctness of the discrete choice (logit) model's specification, the data are consistent with non-separability of the economic incentives and this measure of intrinsic costs of lying. That is, the data confirm that there is heterogeneity within individuals' total costs of lying, again inconsistent with the type-based model's assumption.

Note that $\hat{\beta}_E + \hat{\beta}_{IE}PV_i$ is negative even when evaluated at $PV = 6$. Thus, in the cross-section of participants, the presence of a strong protected value of truthfulness lessened, but did not eliminate, the relevance of the economic costs associated with the earnings management decision. With PV in the regression, the significance of the demographic controls vanishes. It is also noteworthy that, as soon as we include the interaction term with PV , $35HURTS$ is no longer significant.¹⁷ Finally, as shown in Column (5), we also find that our results continue to hold in the subsample without the free-truth situation.

We emphasize that, despite these findings, one cannot conclude that PV has a stronger claim to organizing the data than plausible alternatives. For example, participants may be driven by non-consequentialist preferences that attach expressive utility to low-stakes acts or decisions that substantiate or confirm personal identity. This expressive-preferences concept was developed in the political science literature to explain why citizens vote despite an apparent lack of economic incentive (James M. Buchanan 1954; Gordon Tullock 1971). Expressive preferences have been experimentally documented to play a role in hypothetical choice situations, for example, by Timothy Feddersen, Sean Gailmard, and Alvaro Sandroni (2009). While, in the formulation of these authors, expressive preferences

¹⁷ This is as expected, given the setup of this experiment, and it suggests that the significance of $35HURTS$ in the earlier regressions stems from the fact that this variable (as well as the underlying social-preferences intensity of the individual) is correlated with intrinsic costs of lying. $35HURTS$ has a positive correlation with PV of 0.34. Within our experimental setup, we are unable to address any possible fundamental relationship between protected values and altruistic concerns, so we leave this to future research.

(*EP*) enter utility separably from economic costs, it is conceivable that *EP* also interact with incentives so that non-separability arises. To the extent that the *PV* survey implicitly measures *EP*, and conditional on the correctness of the statistical model’s specification, the results on the interaction term imply that *EP* may, in the range of economic stakes considered in this experiment, become more important in creating a difference in the perceived attractiveness of truthfulness and lying as the stakes increase.¹⁸ Overall, the evidence available from this experiment does not allow us to discriminate definitively between *EP* and *PV* as possible sources of heterogeneity in agents’ lying behaviors. Additional factors relevant in the real world are also not addressed here. For example, intrinsic costs of lying may also be due to an internal reward mechanism for truthfulness that is activated when individuals are, for example, asked to recall the Ten Commandments or to sign an honor code (e.g. Nina Mazar, On Amir, and Dan Ariely 2008).

D. Further results and robustness

Conceivably, participants may have worried about the wealth of the experimenters, which would show up in systematic variation in their choices in the effort task. But, there is no observable relationship between the participants’ levels of effort in that task and their *PV*, *35HURTS*, or socially acceptable responding. This finding also confirms that the experimental design did not simply produce the same pattern of results in the truth-telling and the effort tasks. Moreover, our results are robust to controlling for investment experience, and to variations in samples and estimation methods. All these additional results are available upon request.

V. Conclusion

In this study, we examined individuals who were exposed to a simple, but realistic trade-off: they could tell the truth and suffer economic costs of truthfulness, or they could lie and potentially incur intrinsic costs of lying. In our setting, there was no strategic

¹⁸ Conceptually, *EP* may also be related to tendencies towards self-deception, though the previous results suggest that a standard measure of such tendencies does not explain behavior in this experiment.

incentive to tell the truth; the participants had no counterparty, no notion of a repeated game, no legal obligation, and no risk of being punished.

The experimental results unambiguously reject a type-based model. That is, the results refute Hypothesis TYP, that there exist only “the Ethical” (who care so much about the rightfulness of the process that they always tell the truth) and “the Economic” (who care only about their material payoffs and thus always lie when profitable). Instead, this paper supports Hypothesis HET, reflecting Gneezy’s (2005) conjecture of continuous heterogeneity of preferences for truthfulness: People balance process against consequences in a range of different ways. Moreover, we provide evidence that preferences for truthfulness are non-separable in intrinsic preferences and economic incentives. In sum, our findings point to heterogeneity, both among and within individuals, in their preferences for truthfulness.

This experiment cannot definitively identify the ultimate source(s) of the intrinsic costs of lying. Nor can it state whether the suggested preferences for truthfulness would also be at work at much higher stakes (as the protected-values explanation implies) or whether the validity of the results is limited to relatively low-stakes settings (as the expressive-preferences explanation suggests). Future research may be fruitfully conducted to answer these important questions.

To the extent that preferences for truthfulness apply in a wide range of settings, the results obtained in this study have implications regarding the effectiveness of methods to screen agents for their preferences for truthfulness, as well as implications for the optimal setting of incentive contracts.

References

- Ariely, Dan; Anat Bracha and Stephan Meier.** 2009. "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially " *American Economic Review*, 99(1): 544–55.
- Baiman, Stanley and Barry L. Lewis.** 1989. "An experiment testing the behavioral equivalence of strategically equivalent employment contracts." *Journal of Accounting Research*, 28(1): 1-20.
- Baron, Jonathan and Mark Spranca.** 1997. "Protected Values." *Organizational Behavior and Human Decision Processes*, 70(1): 1-16.
- Bénabou, Roland and Jean Tirole.** 2011. "Identity, Morals, and Taboos: Beliefs as Assets." *Quarterly Journal of Economics*, 126(2): 805-55.
- Bergstresser, Daniel and Thomas Philippon.** 2006. "CEO incentives and earnings management." *Journal of Financial Economics*, 80(3): 511-29.
- Bhide, Amar and Howard H. Stevenson.** 1990. "Why Be Honest if Honesty Doesn't Pay." *Harvard Business Review*, 68(5): 121-29.
- Bowles, Samuel and Sandra Polanía Reyes.** 2011. "Economic Incentives and Social Preferences: Substitutes or Complements?" *Journal of Economic Literature*, forthcoming.
- Buchanan, James M.** 1954. "Individual choice in voting and the market." *Journal of Political Economy*, 62: 334-43.
- Charness, Gary and Martin Dufwenberg.** 2006. "Promises and partnership." *Econometrica*, 74(6): 1579-601.
- Evans, John H.; R. Lynn Hannan; Ranjani Krishnan and Donald V. Moser.** 2001. "Honesty in managerial reporting." *Accounting Review*, 76(4): 537-59.
- Feddersen, Timothy; Sean Gailmard and Alvaro Sandroni.** 2009. "Moral bias in large elections: Theory and experimental evidence." *American Political Science Review*, 103(2): 175-91.
- Fehr, Ernst and Urs Fischbacher.** 2002. "Why social preferences matter." *Economic Journal*, 112(478): C1-C33.
- _____. 2003. "The nature of human altruism." *Nature*, 425(23 October 2003): 785-91.
- Fehr, Ernst and Klaus Schmidt.** 1999. "A theory of fairness, competition, and cooperation." *Quarterly Journal of Economics*, 114(3): 817-68.
- Fischbacher, Urs and Franziska Heusi.** 2008. "Lies in Disguise: An experimental study on cheating." Thurgau Institute of Economics and Department of Economics at the University of Konstanz Research Paper Series 40-2008.
- Gneezy, Uri.** 2005. "Deception: The role of consequences." *American Economic Review*, 95(1): 384-94.
- Grover, Steven L.** 2005. "The truth, the whole truth and nothing but the truth: The causes and management of workplace lying." *Academy of Management Executive*, 19(2): 148-57.
- Healy, Paul M. and James M. Wahlen.** 1999. "A review of the earnings management literature and its implications for standard setting." *Accounting Horizons*, 13(October): 365-83.
- Hurkens, Sjaak and Navin Kartik.** 2009. "Would I lie to you? On social preferences and lying aversion." *Experimental Economics*, 12(2): 180-92.

- King, Gary.** 1998. *Unifying Political Methodology*. Ann Arbor: The University of Michigan Press.
- Koford, Kenneth and Mark Penno.** 1992. "Accounting, principal-agent theory, and self-interested behavior." In *Ethics and Agency Theory: An Introduction*, Norman E. Bowie and R. Edward Freeman, New York: Oxford University Press.
- Mazar, Nina; On Amir and Dan Ariely.** 2008. "The dishonesty of honest people: A theory of self-concept maintenance." *Journal of Marketing Research*, XLV: 633-44.
- Musch, Jochen; Robbi Brockhaus and Arndt Bröder.** 2002. "Ein Inventar zur Erfassung von zwei Faktoren sozialer Erwünschtheit." *Diagnostica*, 48(3): 121-29.
- Nelson, Mark W.; John A. Elliott and Robin L. Tarpley.** 2003. "How are earnings managed? Examples from auditors." *Accounting Horizons*, 17(Supplement): 17-35.
- Paulhus, Delroy.** 1984. "Two-component models of socially desirable responding." *Journal of Personality and Social Psychology*, 46: 598-609.
- Rabin, Matthew.** 1995. "Moral preferences, moral constraints, and self-serving biases." *Unpublished*.
- Rode, Julian.** 2010. "Truth and trust in communication - Experiments on the effect of a competitive context." *Games and Economic Behavior*, 68: 325-38.
- Sánchez-Pagés, Santiago and Marc Vorsatz.** 2007. "An experimental study of truth-telling in a sender-receiver game." *Games and Economic Behavior*, 61: 86-112.
- Sutter, Matthias.** 2009. "Deception through telling the truth?! Experimental evidence from individuals and teams." *Economic Journal*, 119: 47-60.
- Tanner, Carmen; Douglas L. Medin and Rumen Iliev.** 2008. "Influence of deontological versus consequentialist orientations on act choices and framing effects: when principles are more important than consequences." *European Journal of Social Psychology*, 38: 757-69.
- Tanner, Carmen; Bettina Ryf and Martin Hanselmann.** 2009. "Geschützte Werte Skala: Konstruktion und erste Validierung eines Messinstrumentes (Protected Values Measure: Construction and first validation of an instrument to assess protected values)." *Diagnostica*, 55(3): 174-83.
- Tetlock, Philip E.; Orie V. Kristel; S. Beth Elson; Melanie C. Green and Jennifer S. Lerner.** 2000. "The psychology of the unthinkable: Taboo Trade-Offs, Forbidden Base Rates, and Heretical Counterfactuals." *Journal of Personality and Social Psychology*, 78(5): 853-70.
- Treviño, Linda K. and Michael E. Brown.** 2004. "Managing to be ethical: Debunking five business ethics myths." *Academy of Management Executive*, 18(2): 69-81.
- Tullock, Gordon.** 1971. "The Charity of the Uncharitable." *Western Economic Journal*, 9: 379-92.
- Wooldridge, Jeffrey M.** 2006. *Introductory Econometrics: A Modern Approach*. Mason: Thomson Higher Education.
- Zak, Paul J.** 2008. *Moral markets: The critical role of values in the economy*. Princeton and Oxford: Princeton University Press.

TABLES

Table I: Manipulation checks

Participants answered questions that asked for their assessments of announcing 31 and 35 cents, respectively. These questions were on a -2 to +2 scale. After reordering (the direction of the scale varies between questions), a value of +2 indicates that the action was seen as honest, non-manipulative, associated with a personal financial loss, and based on a long-term view.

	31 cents	35 cents	t-test
Honest vs. Dishonest	1.61	-1.17	26.47
Non-manipulative vs. Manipulative	1.39	-1.14	22.70
Personal financial loss vs. Personal financial gain	0.93	-1.25	19.50
Long-term view vs. Short-term view	0.99	-1.1	18.53

Table II: Behavior across economic costs of truthfulness

This table presents the percentages of participants announcing 31 cents of earnings per share across the various economic cost of stating the truth (*ECOST*) conditions.

Percent of participants announcing 31 cents						
<i>ECOST</i> = CHF 0	<i>ECOST</i> = CHF 0.3	<i>ECOST</i> = CHF 0.6	<i>ECOST</i> = CHF 0.9	<i>ECOST</i> = CHF 1.2	Total	Total except <i>ECOST</i> = 0
82.0%	52.1%	31.4%	23.0%	21.1%	41.9%	31.9%

Table III: Determinants of earnings management behavior

This table presents coefficients of logit regressions. The dependent variable is *TRUTHFUL CHOICE*. The explanatory variables are described in the text. Columns (1) to (4) use data from all *ECOST* situations. Column (5) omits the free truth (*ECOST* = CHF 0) situation. Robust standard errors, obtained by clustering at the individual level, appear in parentheses below coefficient estimates. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)
Economic cost of no earnings management (<i>ECOST</i>)	-2.35*** (0.17)	-2.50*** (0.18)	-2.75*** (0.20)	-6.10*** (1.38)	-5.22*** (1.43)
Sex (1: Female, 0: Male)		0.46* (0.26)	0.30 (0.28)	0.30 (0.28)	0.39 (0.30)
Age (Years)		-0.02 (0.03)	-0.03 (0.03)	-0.03 (0.03)	-0.01 (0.03)
Psychology (1: Yes, 0: No)		0.25 (0.29)	0.03 (0.31)	0.05 (0.31)	-0.13 (0.34)
Other studies (1: Yes, 0: No)		0.66* (0.38)	0.24 (0.40)	0.27 (0.40)	0.07 (0.44)
Altruistic concerns (<i>35HURTS</i>)		0.41*** (0.11)	0.22** (0.11)	0.12 (0.11)	-0.04 (0.16)
Impression management tendency (<i>EXTDECEIT</i>)			0.50 (0.97)	-0.07 (0.87)	0.66 (1.13)
Self deception tendency (<i>SELFDECEIT</i>)			0.02 (1.03)	0.08 (0.97)	-1.07 (1.25)
Protected values (<i>PV</i>)			0.73*** (0.14)	0.36*** (0.13)	0.45** (0.19)
<i>35HURTS</i> * <i>ECOST</i>				0.18 (0.21)	0.35 (0.24)
<i>EXTDECEIT</i> * <i>ECOST</i>				1.18 (1.59)	0.14 (1.61)
<i>SELFDECEIT</i> * <i>ECOST</i>				-0.66 (1.74)	0.79 (1.78)
<i>PV</i> * <i>ECOST</i>				0.75*** (0.22)	0.57** (0.25)
Constant	1.00*** (0.11)	0.88 (0.66)	-1.56 (1.16)	0.16 (1.11)	-0.61 (1.26)
Observations	1,305	1,305	1,305	1,305	1,044
Number of participants	261	261	261	261	261
Pseudo R ²	0.15	0.19	0.25	0.26	0.19
Pseudo Log-Likelihood	-757.6	-719.7	-667.4	-655.7	-531.8
Likelihood-ratio test statistic (χ^2 , p-value)	259.6 (<0.01)	335.4 (<0.01)	440.0 (<0.01)	463.4 (<0.01)	243.6 (<0.01)
Wald test statistic (χ^2 , p-value)	183.9 (<0.01)	211.0 (<0.01)	222.8 (<0.01)	204.8 (<0.01)	107.1 (<0.01)

Table IV: Descriptive statistics of important explanatory variables

This table presents descriptive statistics for our measure of altruistic concerns and for three candidate measures of intrinsic costs of lying. N = 261.

	Mean	Median	Std. Dev.	Min	Max
Altruistic concerns (<i>35HURTS</i>)	0.74	1.00	1.07	-2.00	2.00
Tendency towards impression management (<i>EXTDECEIT</i>)	0.49	0.50	0.13	0.16	0.81
Tendency towards self deception (<i>SELFDECEIT</i>)	0.64	0.63	0.12	0.24	1.00
Protected values (<i>PV</i>)	3.82	3.78	1.03	0.00	6.00

Supplementary Appendix

A. 1 Full description of the experiment

PARTICIPANTS

A total of 261 subjects (median age: 23 years) participated in this online experiment. We recruited participants from undergraduate classes at the University of Zurich (Switzerland). 50% of the participants were economics and finance students, 40% psychology students, and 10% students from other fields. 42% were women and 58% were men (distributed across the fields). Participants received a fixed amount as payment for their participation and an additional variable amount as compensation determined by their choices. Anonymity was ensured.¹⁹

PROCEDURE

All participants were first assured of anonymity throughout the experiment, then asked to respond to a few demographic questions and to read some basic instructions. They were informed that they would individually receive a payment, CHF 8, for their completed participation in the study, and an additional payment that depended on their choices.²⁰ The participants then completed the main parts of the experiment: 1) the truthtelling task, 2) the effort task, and 3) the measurement of protected values. (These tasks were not labelled for the participants.) The experiment ended with some final questions serving mainly as control variables; then all the participants were paid. For simplicity, we describe the procedure for only one of the randomized orders of tasks. For both the truthtelling and the effort tasks, participants first were required to demonstrate their understanding of the tasks and of the rules of the experiment.

1) TRUTHTELLING TASK. The truthtelling task consisted of two phases, each involving five choices. Specifically, the task consisted of the following steps: choices in Phase 1, a norm manipulation, choices in Phase 2, and a manipulation check.

Phase 1: In the truthtelling task, each participant was placed in the situation of a CEO who had to announce earnings per share for the previous quarter. The participants were told that the variable component of their salaries depended on the earnings they announced. They were also told that the market currently anticipated the announcement of 35 cents per share as earnings,²¹ but that the true earnings were 31 cents per share. The participants were told that they could announce earnings of 35 cents per share while remaining within legal

¹⁹ Most participants received payment one week after the experiment. For this purpose, each participant had received, before the experiment, a code, based on which the experimenter prepared an envelope containing the earnings. Participants received the sealed envelopes by indicating their personal codes.

²⁰ At the time of the experiment, the exchange rate was about US \$1 = CHF 1.15.

²¹ The actual term for the equivalent of cents in the Swiss currency is “Rappen,” and the experiment used the precise Swiss terminology, that is, a choice between 31 Rappen and 35 Rappen, where 100 Rappen = CHF 1. For simplicity, we refer to “cents” within the text.

accounting limits,²² and that the decision would be solely theirs. They were also informed that they would be paid an amount based on the CEO compensation (according to their choices). This additional experimental payoff was converted into real money at the rate of CHF 100,000 = CHF 0.5. Importantly, participants earned less when choosing to tell the truth.

The participants were then told to submit their financial statements that day. Specifically, they were provided, in a randomized manner, with one of two orders of the following choice tasks:

Which earnings will you announce?

- ☐ 31 cents per share -- In this case, your compensation will be CHF 60,000 (CHF 0.30).
- ☐ 35 cents per share -- In this case, your compensation will be CHF 300,000 (CHF 1.50).

- ☐ 31 cents per share -- In this case, your compensation will be CHF 120,000 (CHF 0.60).
- ☐ 35 cents per share -- In this case, your compensation will be CHF 300,000 (CHF 1.50).

- ☐ 31 cents per share -- In this case, your compensation will be CHF 180,000 (CHF 0.90).
- ☐ 35 cents per share -- In this case, your compensation will be CHF 300,000 (CHF 1.50).

- ☐ 31 cents per share -- In this case, your compensation will be CHF 240,000 (CHF 1.20).
- ☐ 35 cents per share -- In this case, your compensation will be CHF 300,000 (CHF 1.50).

- ☐ 31 cents per share -- In this case, your compensation will be CHF 300,000 (CHF 1.50).
- ☐ 35 cents per share -- In this case, your compensation will be CHF 300,000 (CHF 1.50).

Norm manipulation: After Phase 1, participants were given a page to read stating that there was a good opportunity for the acquisition of another company the following year, for which they would need the shareholders' approval. Participants were randomly assigned to one of three groups: control, positive social norm, or negative social norm. While the control group was given no further information, the positive (negative) social norm group received a social norms manipulation (Robert B. Cialdini, Raymond R. Reno, and Carl A. Kallgren (1990)). It read thus: "One evening, you are sitting with a friend of yours who is a financial analyst. He tells you that increasing reported earnings in order to meet market expectations meets with widespread approval (disapproval)."

Phase 2: After this interlude, all the participants were again provided with the same set of five options as in Phase 1, requiring them to announce earnings of either 31 or 35 cents per share.

Overall, the participants could earn between CHF 9 and CHF 15 in the two phases of the truth-telling task. The participants earned the maximum payment when they engaged in earnings management in both choice situations.

Manipulation check: As our manipulation check, the participants were provided with four items and asked, using a 5-point scale ranging from -2 to +2, the extent to which they judged announcing 31 cents as *dishonest* vs. *honest*, *manipulative* vs. *not manipulative*, and *short-term-oriented* vs. *long-term-oriented*. The same was also done for the 35-cent option. To verify that participants correctly perceived their options, we also asked participants the extent to which they associated announcing 31 cents (or 35 cents) with personal costs or

²² Therefore, risk preferences of individuals did not matter as their choices were not based on the trade-off between the expected benefits and costs of committing a crime.

personal benefits (-2 = associated with personal costs to +2 = associated with personal benefits).

2) *EFFORT TASK*. Participants engaged in a simple calculation task, testing the prediction that protected values of truthfulness would play no role in tasks without an honesty/dishonesty dimension. This also allowed us to examine somewhat the possibility that participants' choices were affected by concerns regarding the experimenter's wealth or by aspects of the experiment's design (for example, the order of choices). In this task, all participants were given the role of a manager who could affect the firm's value and their own remuneration by the amount and the accuracy of the work that they did. Participants were then provided with the following slide:

In this task, you can increase earnings per share and, therefore, your compensation, by working. You will work on five sets of calculations. In each set, you can decide whether to do 1 or 5 simple calculations. Doing 5 calculations takes approximately five times as long as doing 1 calculation, and you will be paid more for this. The compensation you receive for 1 and for 5 calculations will vary over the five sets of calculations. Moreover, you will receive CHF 0.2 for each correct calculation.

Participants were shown an example of a calculation, such as $3 + 4 - 5 + 8 + 3 - 9 = ?$. The participants then read the following screen, one set of choices at a time:

How many calculations do you wish to do?

☐ 1 calculation -- In this case, your compensation will be CHF 60,000 (CHF 0.30).

☐ 5 calculations -- In this case, your compensation will be CHF 300,000 (CHF 1.50).

.... [Three analogous choice situations offering, for 1 calculation, CHF 120,000 (CHF 0.60), CHF 180,000 (CHF 0.90), and CHF 240,000 (CHF 1.20), respectively, are omitted for space reasons.]

☐ 1 calculation -- In this case, your compensation will be CHF 300,000 (CHF 1.50).

☐ 5 calculations -- In this case, your compensation will be CHF 300,000 (CHF 1.50).

On this task, participants could earn between CHF 4.50 (for always choosing one calculation, done incorrectly) and CHF 12.50 (for always choosing five calculations, done correctly).

3) *MEASUREMENT OF PROTECTED VALUES*. According to the correspondence (or compatibility) principle established by Icek Ajzen and Martin Fishbein (1980), values and behavior need to be assessed at a similar level of specificity in order to be able to uncover a link between the two. This principle underlies the measure developed by Carmen Tanner, Bettina Ryf, and Martin Hanselmann (2009), which we used to assess the extent to which participants held truthfulness as a protected value and, therefore, felt committed to truthtelling. Since we are studying earnings management behaviors, we adapted the introduction of their questionnaire to the present context. The questionnaire (see Section II of the Appendix) contains two highly correlated subscales designed to approach protected values from different angles. Five items assessed the participants' reactions to violations of honesty by a hypothetical CEO reporting company information. This represents an indirect approach because the decisions of others were being evaluated. Four additional items assessed the participants' own protected values more directly by examining how much importance they attributed to specific features of protected values (such as trade-off

reluctance, unwillingness to sacrifice a value, or incommensurability), again referring to the specific context of a hypothetical CEO's decisions regarding the reporting of information.

The participants also had to answer another set of questions, which served as control variables. After the experiment, the participants anonymously received their payments of CHF 8 plus their earnings. The average total payment was slightly less than CHF 30.5.²³

DEPENDENT VARIABLES

TRUTHFUL CHOICE. This represented the dependent variable in the truth-telling task, coded as a binary variable that took on the value of 1 if a participant chose to announce earnings of 31 cents (the honest option), and the value of 0 if a participant announced 35 cents (the dishonest option).

EFFORT CHOICE. This represented the dependent variable in the effort task. It took on the value of 1 if a participant chose to do five calculations (high effort), and the value of 0 if a participant chose to do one calculation (low effort).

EXPLANATORY VARIABLES

ECOST. This was a within-participants variation. Costs of truthfulness derived from the amount of money a participant forfeited by announcing 31 cents (that is, by truthfulness). The *ECOST* variable took on values from CHF 0 to CHF 1.20 (= 1.50 – 0.30), in increments of 30 cents.

SOCIAL NORM. This was a between-participants variation. We defined three dummies, making Phase 1 the omitted category in the regressions. *CONTROL* was equal to 1 for all observations from Phase 2 with no additional information, and to 0 otherwise. *POSITIVE NORM* was equal to 1 for all observations from Phase 2 with the positive social norm (approval of earnings management), and to 0 otherwise. *NEGATIVE NORM* was equal to 1 for all observations from Phase 2 with the negative social norm (disapproval of earnings management), and to 0 otherwise.

PROTECTED VALUES (PV). After appropriate recoding of some items, an index of the degree of protected values (*PV*) was constructed, based on the means across the four direct and the five indirect items. This index took on a value between 0 (for an individual with no protected values) and 6 (for an individual with maximum protected values). The internal consistency of this scale, as assessed by Cronbach's α , was very satisfactory ($\alpha = 0.86$).²⁴

²³ As explained earlier, by using codes to distribute earnings, we took as much care as possible to ensure anonymity. That is, we tried to remove any possible grounds for expecting reciprocity. It is, therefore, unlikely that a desire to appear honest affected the participants' behavior systematically. See Dan Ariely, Anat Bracha, and Stephan Meier (2009) for a study of how publicly displayed monetary incentives can be less effective in promoting pro-social behavior than privately displayed incentives.

²⁴ Cronbach's Alpha is a measure of the reliability and the internal consistency of an instrument. The measure ranges from 0 to 1 and will generally increase when the correlations between the items increase. We also did the analysis using the direct and indirect subscales separately, with similar results.

CONTROL VARIABLES

ALTRUISTIC CONCERNS. We asked participants the extent to which they believed that announcing 31 cents (or 35 cents) had consequences for other stakeholders (-2 = *hurting other stakeholders* to +2 = *not hurting other stakeholders*). Of course, within the strict confines of the experiment, there were no such consequences. Nonetheless, this variable was a potentially relevant control for altruistic preferences or fairness concerns that participants might have and that might confound our inferences regarding protected values of truthfulness. Answers to this question were coded as the variable *35HURTS*.

SOCIALLY ACCEPTABLE RESPONDING. We used the standard Deception Scales (PDS) of Delroy Paulhus (1984); see Jochen Musch, Robbi Brockhaus, and Arndt Bröder (2002) for the German version. This questionnaire measured individuals' tendencies to give socially desirable responses (SDR), in two distinct forms: self-deception and impression management. Accordingly, we coded two variables *SELFDECEIT* and *EXTDECEIT*. Participants who exhibited more socially acceptable responses scored higher on both scales.

DEMOGRAPHIC CONTROL VARIABLES. *SEX*, *AGE*, and *STUDIES* served as control variables. *SEX* was equal to 1 for female participants and to 0 for male participants. *AGE* was equal to each participant's age in completed years. *PSYCHOLOGY* was equal to 1 for psychology students ("psychologists") and to 0 otherwise. *OTHER* was equal to 1 for participants from fields other than psychology and economics and to 0 otherwise. *ECONOMICS* was the omitted category.

A. 2 Protected Values survey

According to the correspondence (or compatibility) principle established by Ajzen and Fishbein (1980), values and behavior need to be assessed at a similar level of specificity in order to be able to uncover a link between the two. This principle underlies the measure developed by Tanner, Ryt, and Hanselmann (2009).

Since we are studying earnings management behaviors, we adapted the introduction of their questionnaire to the present context. The questionnaire, contains two highly correlated subscales designed to approach protected values from different angles.

Five items assessed the participants' reactions to violations of honesty by a hypothetical CEO reporting company information. This represents an indirect approach because the decisions of others were being evaluated.

Four additional items assessed the participants' own protected values more directly by examining how much importance they attributed to specific features of protected values (such as trade-off reluctance, unwillingness to sacrifice a value, or incommensurability), again referring to the specific context of a hypothetical CEO's decisions regarding the reporting of information.

After appropriate recoding of some items, an index of the degree of protected values (PV) was constructed, based on the means across the four direct and the five indirect items.

Note: The original Protected Values survey was conducted in German. In the paper, for ease of interpretation of the empirical results, we changed the scale to range from 0 to 6.

Because CEOs' compensation levels depend on the earnings they report to their shareholders, CEOs have an incentive to modify reports to shareholders. What is your opinion on CEOs modifying company information in reports?

Please choose the appropriate category. This is:

Very immoral	1	2	3	4	5	6	7	very moral
Not at all praiseworthy	1	2	3	4	5	6	7	very praiseworthy
Not at all blameworthy	1	2	3	4	5	6	7	very blameworthy
Not at all outrageous	1	2	3	4	5	6	7	very outrageous
Not at all acceptable	1	2	3	4	5	6	7	very acceptable

CEOs have an opportunity to modify information in the reports they provide to their shareholders. Some view such modification as a violation of truthfulness, others regard it as acceptable protection of personal interests. What do you think about the value of truthfulness in such a situation?

Truthfulness is something

... that one should not sacrifice, no matter what the (material or other) benefits.								
strongly disagree	1	2	3	4	5	6	7	strongly agree
... for which I think it is right to make a cost-benefit analysis.								
strongly disagree	1	2	3	4	5	6	7	strongly agree
... that cannot be measured in monetary terms.								
strongly disagree	1	2	3	4	5	6	7	strongly agree
... about which I can be flexible if the situation demands it.								
strongly disagree	1	2	3	4	5	6	7	strongly agree

A. 3 Additional results

A. Distribution of PV

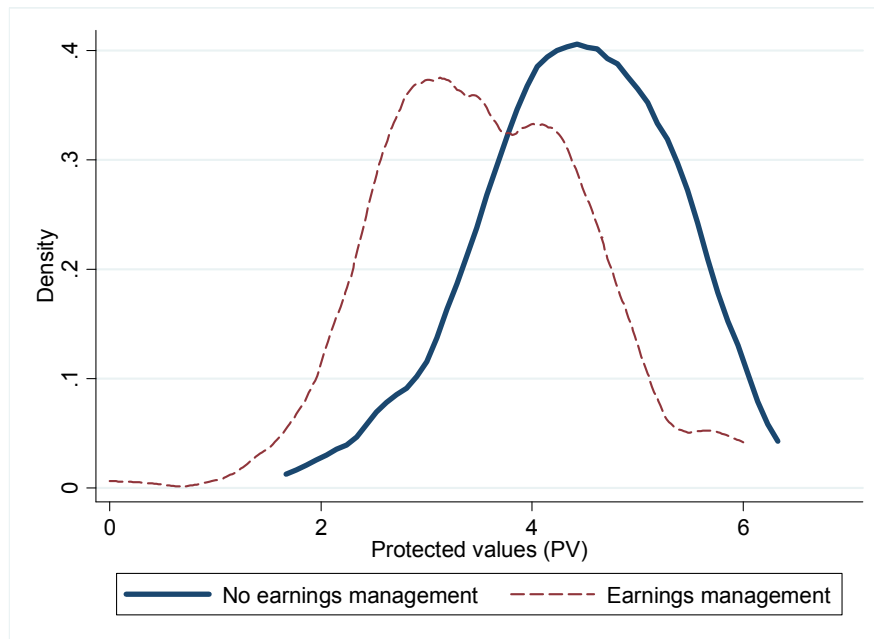


Figure S-1: This figure plots the kernel density (using an Epanechnikov kernel) of the strength of protected values of the participants who chose to manage earnings (broken line) and for those participants who willingly bore economic costs for not managing earnings (solid line). The figure is plotted for the median cost level in Phase 1. Protected Values (PV) is the average of all nine items of the survey described in the text. Higher numbers correspond to stronger protected values.

B. Quantitative implications for preference parameters

In the main text, we found that PV serves as a useful measure to organize the data. Here, we provide a quantitative interpretation of the findings for PV . (The other candidate measures for intrinsic costs of lying were insignificant.) For PV , a useful way to interpret the results quantitatively is offered by slightly rewriting the preference specification in equation (1). Let $V_i(T | PV_i) = [b(m - ECOST) + bECOST(1 - k_i)(1 - T)] - \ell_i(1 - T)$. The difference between the utilities of truthtelling and of lying is here given by

$$Y_i^* = \ell_i - bECOST(1 - k_i). \quad (A1)$$

In this formulation, the role of intrinsic costs of lying is split up even more explicitly than in the formulation in the main paper, although they are mathematically isomorphic. Here, first, protected values generate moral costs of lying $\ell(PV_i) = \ell_i$. A second role of protected values is that the actual marginal utility an agent ascribes to monies depends on the process by which the funds were obtained and on the agent's moral evaluation of that process. In the present context, a dollar obtained by an act of lying may be regarded as less valuable than a dollar obtained by telling the truth. This is reflected in a preference parameter $k(PV_i) = k_i$. An agent discounts funds obtained by lying if k_i is positive.

We posit for parsimony the simple linear parametric specifications $\ell_i = \zeta_0 + \zeta_1 PV_i$ and $k_i = \eta_0 + \eta_1 PV_i$. Adding a stochastic error and rearranging terms, each participant i 's ($i=1, \dots, 261$) latent utility difference between truthtelling and lying at direct economic $ECOST_j$ is then given by

$$\begin{aligned} Y_{ij}^* &= \ell_i - bECOST_j(1 - k_i) + \varepsilon_{ij} = \\ &= \zeta_0 + \zeta_1 PV_i - (1 - \eta_0)bECOST_j + \eta_1 bPV_i ECOST_j + \varepsilon_{ij}. \end{aligned} \quad (A2)$$

Thus, under the maintained distributional assumptions, after relabeling and combining coefficients,

$$\Pr(T_{ij} = 1 | \mathbf{X}) = \Lambda[\phi_0 + \phi_{PV} PV_i + \phi_{ECOST} ECOST_j + \phi_{PVECOST} PV_i ECOST_j], \quad (A3)$$

where $\Lambda(\bullet)$ is the logistic cumulative distribution function. For the model estimated in Column (1) of Table III of the paper, we therefore have $\hat{\ell}_i = \hat{\phi}_0 + \hat{\phi}_{PV} PV_i$ and $\hat{k}_i = 1 + (\hat{\phi}_{ECOST} + \hat{\phi}_{PVECOST} PV_i)(1/b)$. For k_i , the parameters η_0 and η_1 are not identified, because b is unknown. One way to make progress is to determine (speculatively) k_i at some PV level. As a perhaps reasonable benchmark, assume that k is zero for an individual with an average PV level (3.82). From this, we can infer the implied $b = 2.91$. (Recall that we are assuming that b is independent of protected values.) This in turn then allows us to plug in a range of PV levels to obtain the implied lying discounts (or premia). (Another approach is to simply posit values for b and calculate k_i accordingly.)

Table S-I shows the results of these calculations. 95% confidence intervals are shown in brackets below the point estimates. For the moral cost of lying, ℓ_i , the estimates suggest an average value of around unity. If the average participant is indifferent to the process by

which funds are obtained, our estimate of b is 2.91. For the discount parameter, k_i , our experiment then implies that an individual with a protected value in the 75th percentile will discount funds acquired dishonestly by about 21%, with even the lower bound of the 95% confidence interval being about 8%. Thus, the trade-off resistance implied by protected values is not only a statistically, but also an economically, significant factor in decision making.²⁵ Of course, assuming that the average k_i is zero implies that an economic model using the proposed preference specification becomes interesting mostly when there is variation in protected values across agents.

Table S-I: Quantitative implications for preference parameters

	<i>Average PV</i>	<i>PV at 25th percentile</i>	<i>PV at 75th percentile</i>	<i>PV=6</i>
t_i	1.15 [0.93; 1.37]	0.79 [0.52; 1.07]	1.46 [1.18; 1.75]	2.09 [1.54; 2.64]
k_i	0 n.a.	-0.23 [-0.41; -0.04]	0.21 [0.08; 0.35]	0.63 [0.36; 0.89]

²⁵ Note that factors reflected neither in the model nor in the experiment may shift the attractiveness of truthfulness. For example, an anticipated loss of reputation in case of cheating may make lying less attractive at a given cost of truthfulness, thus adding to the personal moral cost of lying. These external factors would, therefore, complement and enhance the power of the discount factor k_i .

Appendix references

- Ajzen, Icek and Martin Fishbein.** 1980. *Understanding Attitudes and Predicting Social Behavior*. Englewood Cliffs: Prentice-Hall.
- Ariely, Dan; Anat Bracha and Stephan Meier.** 2009. "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially " *American Economic Review*, 99(1): 544–55.
- Cialdini, Robert B.; Raymond R. Reno and Carl A. Kallgren.** 1990. "A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places." *Journal of Personality and Social Psychology*, 58: 1015-26.
- Musch, Jochen; Robbi Brockhaus and Arndt Bröder.** 2002. "Ein Inventar zur Erfassung von zwei Faktoren sozialer Erwünschtheit." *Diagnostica*, 48(3): 121-29.
- Paulhus, Delroy.** 1984. "Two-component models of socially desirable responding." *Journal of Personality and Social Psychology*, 46: 598-609.
- Tanner, Carmen; Bettina Ryf and Martin Hanselmann.** 2009. "Geschützte Werte Skala: Konstruktion und erste Validierung eines Messinstrumentes (Protected Values Measure: Construction and first validation of an instrument to assess protected values)." *Diagnostica*, 55(3): 174-83.

swiss:finance:institute

c/o University of Geneva
40 bd du Pont d'Arve
1211 Geneva 4
Switzerland

T +41 22 379 84 71
F +41 22 379 82 77
RPS@sfi.ch
www.SwissFinanceInstitute.ch